

Task 3 – Relationship of the GeoLibrary to the proposed GeoArchives

Overview Comment

From the point of view of the GeoArchives, close integration with the developing operations and evolution of the GeoLibrary is vitally important. In a sense the GeoArchives can be viewed as a ‘wing’ of the GeoLibrary, much as many traditional archives are integrated with a library. In fact, it is possible that the data sets transitioning to GeoArchives status could stay on the same servers and just be reappraised and reclassified with new metadata and given different access protocols.

There are numerous reasons for this close coordination, and many are discussed throughout the Workplan and other project documents, but basically there is great overlap of numerous aspects of the goals and mandates of each of these institutions. Efficiency will be greatly increased with tight coordination including:

- Data sets- Most but not all data sets will come into the GeoArchives from the GeoLibrary. Interagency coordination will facilitate efficient data transfer to the GeoArchives
- Metadata requirements- GeoLibrary FGDC metadata will be the foundation of the GeoArchives metadata. Some additional fields not in FGDC may need to be added from the Dublin Core , but it should be accessible as one metadata system
- Software and hardware access tools- The best use of GIS data sets require GIS software access tools
- User (client) base- users of the Maine GeoLibrary will often be the same constituency that requests and accesses historic GeoArchives data sets.
- Staff knowledge requirements- GIS expertise will be continually needed by GeoArchives staff to serve the user community
- Data export requirements – This is an example of needed GIS expertise in the GeoArchives
- Data migration and Technical Assistance- GeoLibrary staff will have ongoing technical training that will be of benefit as the GIS state-of-the-art methodologies evolve.
- Clients of the GeoArchives will often be using GeoLibrary data sets in conjunction with GeoArchives layers
- The GeoLibrary staff will have good knowledge of data sets, and this information will supplement the standard metadata as the GeoArchives determines methods and procedures for adding data sets to the GeoArchives
- Archives staff will be invaluable in advising GeoLibrary staff on numerous issues including, appraisal, metadata fields and attribute population.

Relationship of GeoArchives the GeoLibrary

Joint planning and operations between these two organizations (GeoLibrary and GeoArchives) are correctly recognized as essential for the success of both programs. Therefore it is important to have both a good working relationship and a good formal

relationship between the GeoLibrary and the planned GeoArchives. The workplan for the GeoArchives taps the GeoLibrary Board as the advisory committee to the GeoArchives project.

However the Legislation creating the Maine GeoLibrary does not explicitly define a role for the Maine Archives on its controlling Geographic Advisory Board or other capacity. The staff and operations of the Maine Archives have expertise and experience to assist the GeoLibrary in its startup operations, especially planning the integration of the metadata standards- FGDC for the GeoLibrary, and the Dublin core for the GeoArchives. The planned formal SLA agreements between the 2 groups described in the Workplan are essential and adequate to meet this and other needs.

Implications for Web Services Data vs. Data Stored on GeoLibrary Servers

One of the major decisions/negotiations to be made during the development of the GeoLibrary is how much of the data is actually stored on GeoLibrary servers and how much is accessed via web services and therefore stored on other servers managed by other institutions.

While web services connected to an institution such as USGS will be relatively stable and reliable over time for the GeoLibrary, institutional agreements will need to be made for the data sets to be considered secure for the GeoArchives. Alternatively the data sets and metadata will need to be directly integrated into the Maine GeoArchives and therefore under the complete control and management of the GeoArchives.

Smaller institutions that host data sets used by the GeoLibrary will definitely not be suitable for long-term cooperative hosting agreements, and these data, if judged worthy of inclusion in the GeoArchives, will need to be transferred into the GeoArchives along with their metadata.

If the task of transferring these data onto its servers is taken on by the GeoLibrary during the data sets 'active' life, there should be little or no additional work for the transferring this data to the GeoArchives as compared to importing and standardizing other data sets in the GeoLibrary. If the GeoLibrary accesses the data via web services, then there will be additional overhead and work to bring qualifying data into the GeoArchives and process it for permanent retention in the GeoArchives.

Electronic Standards & the unique nature of geographic data

The GeoArchives should adopt existing electronic archives standards and methods where possible, especially electronic database archival standards, as described in the Maine State Archives Digital Records Management Plan. But also be aware of the special nature of geography and geographic data sets and how they are used and analyzed. Budget for and be prepared to modify the format of these GIS data as the 'state-of-the-art' standards for both databases and GIS data evolve, as they surely will.

Preservation and Access

The migration (conservation) of digital data sets to various future formats must be carefully planned and done with consideration of the two pillars of archival theory - preservation of the data and access by users.

Preservation

A major long-term trend in GIS has been the evolution of GIS software and data towards de-facto information technology (IT) standards, e.g. Oracle ®. This trend will be of great utility to archivists, because they will presumably be knowledgeable about these IT standards in digital data preservation and management from their work with the many non-geographic data sets of the State.

Access

The most difficult part of the GeoArchives for non-GIS-specialist archivists will be the access issues, because of the complex nature of geography and the unique ways that disparate sets of geographic data can and should be used; e.g. spatial and overlay analysis between data sets; aggregation and generalization; change over time; and the great power of data visualization. This means that GIS data is best accessed with the tools and within the context of modern and evolving GIS software and applications.

For this reason as well as others noted in this discussion it is in the best interest of the Archives to have a close and long-term formal relationship the Maine GeoLibrary and Maine operational GIS organizations such as the Office of GIS, who presumably will always maintain current knowledge of methods and trends and can therefore assist in data migration planning as well as vetting GIS data sets for potential inclusion in the GeoArchives.

Metadata

The GIS community generally relies on the FGDC metadata standards. The Archives world has many standards, but the Dublin Core is proposed in the Functional Requirements document for the GeoArchives. Virtually all of the metadata called for in the Dublin Core is already present on the FGDC standards. Therefore it may be more efficient for the GeoArchives to simply adopt the broader FGDC standards. If there needs to be a Dublin Core subset extracted, simply mapping Dublin Core fields to the FGDC template would be more efficient than reformatting and duplicating these fields in the more familiar Dublin Core.

The GeoArchives should be aware that GIS programs and professionals often do not have the time and resources to enforce full metadata standards for all data sets they create and/or collect. Enforcement of these standards will be a continuing challenge to the GeoLibrary and subsequently the GeoArchives.

Types of GIS data sets

(this section is copied from my commentary on the GIS staff's technical recommendation notes document):

There are 3 different types of georeferenced data sets. The GeoArchives must efficiently and consciously accommodate all of these types of updated data sets in the best possible manner:

- Raster data, generally imagery, where there is essentially one continuous data set for the state or subunit of the state. These data are not feature-based nor updated incrementally.
- Vector data (point;line;polygon;route, etc.) where the GIS data is continually updated at the feature level (e.g. E911). This type of data needs metadata at the feature level to capture information on when and how and when individual feature data was collected.
- Vector data (point;line;polygon;route) where the data is updated for the whole data set at discrete, periodic intervals (e.g. possibly public wells).

NOTE: Vector data is further complicated by the fact that some data sets updated periodically have attributes that are updated much more frequently (e.g. temporal measurements of water levels or radon levels in wells). As noted below by OGIS staff, ArcSDE is a good step in the right direction for feature-based metadata, but it does not capture attribute edits and changes transactionally. Therefore special attention must be paid if the archives decided it is important to maintain the history of attribute modification or change. One solution is to have multiple fields used to contain the history of an changing attribute.

Retaining and Archiving Types of Data Sets

- Retain all copies of periodically updated data sets, e.g. orthophotography or vector data that is updated on a regular schedule. Do not take 'snapshots' at regular intervals, just when the data set undergoes an upgrade or complete replacement. An imagery data set should be archived whenever a successor data set is loaded into the GeoLibrary and the old one is no longer the data of choice for general GIS use.
- Take 'snapshots' of continually edited data sets at regular intervals. The time between snapshots should be decided on a case-by-case basis. Even though it is theoretically possible to have one data set (roads is used in the GeoArchives workplan) with 50 years of history that one can manipulate to go back to the roads as they were several years before, it is advisable to make a discrete archival copy of such a data set at periodic intervals. In this way, if the data becomes corrupted, there is a backup of available.
- The GeoLibrary should be encouraged to track the usage of data through the GeoLibrary. This will inform the evaluation of data sets by the GeoArchives. But to reiterate, it is expected that most all data that meets the criteria of the GeoLibrary will be worthy of migration to and inclusion in the GeoArchives when its currency for the GeoLibrary is past.

Security

The use of standard computer IT based strategies for maintaining secure access to the GeoArchives is recommended, and this should be adequate. Group and individual passwords are an example of this. I envision no additional issues of security caused by the geographic nature of these databases and data sets.

Privacy

Generally the department that creates the GIS record is responsible for dealing with any legal or moral privacy issues, just as it is for any other public record. GIS data sets that are processed by and located in the GeoLibrary will have been filtered for privacy issues, e.g. name or specific address of a cancer patient. The Maine Archives staff may have expertise in this area that could be of benefit to GeoLibrary staff.

Feature metadata

Feature metadata has been discussed by the GIS Technical Group Feature Data Subcommittee (2000), and it included in the Workplan for the GeoArchives. An attribute should be added when a feature is added, updated, corrected or deleted from a data set in the GeoLibrary. It will be important to have this ability, especially for data sets updated continually, as it would otherwise be difficult to see the full history of a specific feature. The OGIS staff plans to test this as part of their beta ArcSDE 9.1 project. This is a promising, if proprietary, GIS software advance and the Archives should monitor the progress of this test. In general, when the data sets are transferred to the GeoArchives, feature metadata will be frozen.

Deaccession

Develop the rules and the process by which to delete features or data sets in the future. What are the current Maine Archives standards for deaccessioning data in other record series? I did not see it mentioned in the Maine State Archives “Digital Records Management Plan”. Rules for other digital records will probably be adaptable for GIS data sets.

Other Archives Series with geographical location attributes

The Archives, as it plans the GeoArchives, should keep in mind that much information currently in the Archives, is not explicitly considered geographic but does have a geographic component. For example a database of Maine businesses will almost always have a geographic location (e.g. the address of each facility.). The Archives must be prepared to respond to such requests to merge GeoLibrary data with other archives records.

In this sense, the GeoArchives has the potential to be a very useful ‘front-end’ for some of their other series and collections. In some cases this will require the relatively simple generation of location points from address fields in data bases. Other potential instances are:

- the scanning and geo rectification of hardcopy maps to an appropriate GIS base map; and
- the digitization of paper records and manual creation of addresses and/or other locational

While it would not be generally expected that Maine Archives staff actually plan and perform this type of work for Archives users, staff will need to be able to assist the researcher understand the nature of what it would take to perform that work and point them in the right direction.